



APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

Report Documentation Page		Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE 27 SEP 2010	2. REPORT TYPE Monograph	3. DATES COVERED 01-01-2008 to 01-12-2008
4. TITLE AND SUBTITLE Initial Validation of Novel Performance-Based Measures: Mental Rotation and Psychomotor Ability		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Philip Fatolitis; Florian Jentsch; Peter Hancock; Robert Kennedy; Clint Bowers		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Aerospace Medical Research Laboratory, 280 Fred Bauer Street, Building 1811, Pensacola, FL, 32508		8. PERFORMING ORGANIZATION REPORT NUMBER NAMRL 10-52
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Navy Medicine Manpower, Personnel, Training, & Education Command, 8901 Wisconsin Avenue, Bethesda, MD, 20889		10. SPONSOR/MONITOR'S ACRONYM(S) NAVMED MPT&E
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT Given the high-risk nature of military flight operations and the significant resources required to train U.S. Naval Aviators and Flight Officers, personnel selection must continually be improved. In addition to general commissioning requirements and aeromedical standards, the U.S. Navy utilizes the Aviation Selection Test Battery (ASTB) to select commissioned aviation students. Although the ASTB has historically proven to be a good predictor of aviation student performance in training, it is proposed here that incremental improvement may be gained with the introduction of novel, computer administered performance-based measures: the Block Rotation Task (BRT) and a Navy-developed 2-D compensatory tracking task. This work constituted an initial validation of the two tasks. The BRT is an interactive virtual analog of Shepard-Metzler's (1971) mental rotation task. The 2-D tracking task is typical of what is found in the literature, but was developed by the U.S. Navy. The BRT was investigated for its suitability as a measure for quantifying both mental rotation and psychomotor ability. Data from the BRT were examined to determine task reliability and to formulate a quantitative/predictive performance model of combined human mental rotation and psychomotor ability. Data gathered from the compensatory tracking task were investigated to determine if they concord with results in extant literature, indicating the validity of the task. Results showed that both the BRT and the 2-D tracking tasks are measures of spatial ability. A descriptive performance model of the BRT is also presented.		
15. SUBJECT TERMS Block Rotation Task, aviation personnel selection, 2-D compensatory tracking task		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 47	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Reviewed and Approved
27-09-2010



Rita G. Simmons, CDR MSC USN



The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

This work was completed during Full Time Our Service Training (FTOST) supported by the Navy.

The study protocol was completed in compliance with all applicable Federal regulations governing the protection of human subjects.

I am a member of the U.S. Government . This work was prepared during FTOST. Title 17 U.S.C. §105 provides that 'Copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

CONTENTS

LIST OF FIGURES.....	v
LIST OF TABLES.....	vi
CHAPTER 1: INTRODUCTION.....	1
Problem Statement.....	1
Research Goals.....	1
CHAPTER 2: LITERATURE REVIEW.....	3
Spatial Ability.....	3
Operational Definition.....	3
Associated Neurological Processes.....	3
High Level Definitions of Spatial Ability Constructs.....	4
An Ecological Explanation for SpA Constructs.....	5
Gender Differences.....	5
Quantification of Spatial Ability.....	6
Mental Rotation.....	9
Psychomotor Ability.....	11
Operational Definition.....	11
Associated Neurological Processes.....	11
Application of Psychomotor Data.....	11
Computer-Based Medium.....	12
Summary.....	12
CHAPTER 3: METHOD.....	14
Participants.....	14
Research Design.....	14
Materials.....	15
Pre-Simulation Questionnaire.....	15
Spatial Apperception.....	15
Block Rotation Task.....	15
Vandenberg Mental Rotation Test.....	16
Manikin Test.....	16
2-D Compensatory Tracking.....	17
Procedure.....	17
CHAPTER 4: RESULTS.....	18
Data Screening.....	18
Descriptive Statistics.....	18
Block Rotation Data Screening and Cleaning.....	18
Assessment Reliability.....	23
Approach and Analysis.....	23
Gender Differences.....	23
Hypothesis I.....	24
Hypothesis II.....	29
Hypothesis III.....	30
CHAPTER 5: DISCUSSION AND CONCLUSION.....	32
Initial Remarks.....	32
Hypothesis I.....	32

Hypothesis II.....	33
General Discussion.....	33
Gender.....	33
Hypothesis III.....	33
Time and Performance.....	34
Study Limitations.....	35
Recommendations.....	35
Conclusion.....	36
REFERENCES.....	37

LIST OF FIGURES

FIGURE 1.....	7
FIGURE 2.....	8
FIGURE 3.....	8
FIGURE 4.....	10
FIGURE 5.....	10
FIGURE 6.....	16
FIGURE 7.....	20
FIGURE 8.....	21
FIGURE 9.....	22
FIGURE 10.....	31
FIGURE 11.....	31

LIST OF TABLES

TABLE 1.....	15
TABLE 2.....	19
TABLE 3.....	19
TABLE 4.....	25
TABLE 5.....	26
TABLE 6.....	27
TABLE 7.....	28
TABLE 8.....	29
TABLE 9.....	29
TABLE 10.....	29
TABLE 11.....	30

ACKNOWLEDGMENTS

This monograph is the published version of dissertation work completed at the University of Central Florida in 2008. The time, effort and talent of the author's advisors, Florian Jentsch, Ph.D., Peter Hancock, D.Sc., Ph.D., Robert Kennedy, Ph.D., and Clint Bowers, Ph.D. are very much appreciated.

ABSTRACT

Given the high-risk nature of military flight operations and the significant resources required to train U.S. Naval Aviators and Flight Officers, personnel selection must continually be improved. In addition to general commissioning requirements and aeromedical standards, the U.S. Navy utilizes the Aviation Selection Test Battery (ASTB) to select commissioned aviation students. Although the ASTB has historically proven to be a good predictor of aviation student performance in training, it is proposed here that incremental improvement may be gained with the introduction of novel, computer administered performance-based measures: the Block Rotation Task (BRT) and a Navy-developed 2-D compensatory tracking task. This work constituted an initial validation of the two tasks. The BRT is an interactive virtual analog of Shepard-Metzler's (1971) mental rotation task. The 2-D tracking task is typical of what is found in the literature, but was developed by the U.S. Navy. The BRT was investigated for its suitability as a measure for quantifying both mental rotation and psychomotor ability. Data from the BRT were examined to determine task reliability and to formulate a quantitative/predictive performance model of combined human mental rotation and psychomotor ability. Data gathered from the compensatory tracking task were investigated to determine if they concord with results in extant literature, indicating the validity of the task. Results showed that both the BRT and the 2-D tracking tasks are measures of spatial ability. A descriptive performance model of the BRT is also presented.

CHAPTER 1: INTRODUCTION

Problem Statement

Notwithstanding the high-risk nature of U.S. Naval Aviation training and operations, the Navy spends approximately \$1 million to train each Naval Aviator and Naval Flight Officer from the time of accession to the time of Advanced Flight Training completion (about 3 years). This substantial cost underlies the necessity of a rigorous personnel selection process. The U.S. Navy's current Aviation Selection Test Battery (ASTB) is predictive of aviation training outcomes. In a recent validation study, it was found that ASTB composite scores had the following values for predicting final grades upon completion of U.S. Navy Primary Flight Training: Academic Qualification Rating $r = .45$ ($p < .001$), and Pilot/Flight Officer Flight Aptitude Rating $r = .35$ ($p < .001$) (ASTB, 2006). Although these correlational values are reasonable, an incentive exists to strive for continued improvement. It is estimated that \$6M (million) training dollars per year or more can be saved for each 5 percentage points of variance in training performance accounted for by means of personnel selection (U.S. Navy Flight Surgeon's Manual, 1989).

Among cognitive tests in a wide range of talents, general intelligence/ability accounts for roughly 50% of common variance, while quantitative, spatial, and verbal ability each account for approximately 8%-10% of the remaining common variance (Lubinski, 2004). It follows that if even 1% of common variance can be accounted for via the increased capacity to quantify relevant aptitude and/or ability, the U.S. Navy could save \$1M training dollars or more per year. This provides the rationale to investigate additional forms of assessments for U.S. Navy Aviation personnel selection.

Although the work presented here did not investigate the predictive validity of the tasks, the research may show whether the tasks under investigation possess psychometric properties that are desirable for further study. In particular, the present work focused on determining the reliability and incremental validity of performance-based measures (PBMs) consisting of spatial and psychomotor ability assessments, as these assessments appear to offer an attractive source of new approaches to personnel testing.

Research Goals

This research constituted an initial validation effort with regard to two novel computer-based PBMs: the Block Rotation Task (BRT) and (2-D) Compensatory Tracking. The goal of this study was to determine if the novel tasks indicate potential utility for aviation personnel selection in the U.S. Navy.

The BRT is a derivative of the Shepard-Metzler (1971) mental rotation task. It consists of a set of virtual 3-D blocks: one being the target stimulus and the other being the comparison figure. Participants' goal in this task is to manipulate the comparison figure, in 3 dimensions, into a matching orientation to the target figure as quickly as possible. The Compensatory Tracking task is the U.S. Navy's version of a computer-

based task that has a strong presence in the literature. This task requires participants to keep a cursor in the center of a monitor screen even though the cursor moves in a variable fashion and the participant has limited control, in 2 dimensions, over the behavior of the cursor.

Both tasks possess relatively short administration times, with no more than 20 minutes for BRT, and one minute or less for Compensatory Tracking. BRT was hypothesized to make demands on mental rotation (a form of spatial ability), psychomotor ability, and coordination. Compensatory Tracking is thought to make more of a demand on “pure” psychomotor ability and relatively less on spatial processing.

These tasks are being considered for operational use due to the fact that PBMs have historically proven to be good predictors of performance in flight training and that any increase in the predictive power of personnel selection tests yields exponential returns in training dollars. Early in the 20th century, analog devices were used for aviation personnel selection but were discontinued because they were not easily “co-calibrated.” Further, the devices were located in a few, isolated stations, a situation that conflicted with evolving recruiting practices. However, given the ubiquity and power of modern computing devices, the constructs that were previously measured with analog devices can now be measured with reliable, easily co-calibrated digital computing devices.

This experimental approach examined various potential measurement capabilities and features of the BRT. This was achieved primarily via comparison of performance on the BRT to that of other validated performance-based measures. The hypotheses of this study were:

- I) BRT correlates with both psychomotor and mental rotation tests while psychomotor and mental rotation tests do not correlate.
- II) Validated performance tasks express differential predictive models for the BRT in an hierarchical regression.
- III) BRT normative performance expresses a linear chronometric quantitative model similar to that found in Shepard-Metzler (1971).

Although normative data were sought to confirm that the novel compensatory tracking task concords with the behavior of other tracking tasks in the literature, it was assumed that the novel tracking task’s measurement characteristics were similar to extant tasks. Normative and reliability data were examined for the BRT with the expectation that error reduction latency increased with increased angular disparity. The intention of this investigation was to set the foundation for further validation work given confirmation of hypotheses or other indication that further investigation is warranted.

CHAPTER 2: LITERATURE REVIEW

Spatial Ability Operational Definition

Lohman (2000) indicated that like verbal ability, spatial ability (SpA [distinguished from the oft-used abbreviation for situational awareness, SA]) is a form of general fluid ability (intelligence), expressed in the form of inductive reasoning as spatial processing occurs. As a stand-alone construct, *SpA can be operationally defined as “the ability to generate, retain, retrieve, and transform well-structured visual images whose properties include location, size, distance, direction, separation, connection, shape, pattern, and movement” (Lohman, 1993, p. 1).* Lohman’s definition is widely adopted in the literature and has served as the working definition of SpA for many research projects.

Not unlike other measures of cognitive ability, properly quantifying SpA performance includes explaining systematic individual differences that are uniquely spatial, and defining the portion of variation on spatial tasks that is shared with more general abilities (Sternberg, 2000). Sternberg’s approach to parsing variability is applied in this research in the sense that shared variance among selected spatial and psychomotor tasks is examined.

Associated Neurological Processes

Neurologically, the processing of spatial information is fundamentally distributed throughout the visual-perceptual system. Spatial processing is traditionally considered to be a right hemispheric phenomenon given the dominance in left hemispheric structures for processing linguistic material.

However, experimental findings have shown that spatial processing, although residing to a significant degree in the right hemisphere, can be expressed in a distributed fashion. After the bifurcation of the early visual structures into ventral and dorsal streams in the lateral geniculate nucleus, some spatial processing occurs in the inferior temporal area, but structures in the posterior parietal area are closely associated with spatial processing proper (Kandel & Wurtz, 2000). Additionally, Kandel, Kupferman, and Iverson (2000) presented evidence that processing spatial material from memory produces activation in the right hippocampal, parahippocampal, and parietal areas. Despite studies that continue to corroborate specialization in the right hemisphere for spatial processing, Levin, Mohamed, and Platek (2005) also implicated the left parahippocampal gyrus, areas of the frontal gyrus, and other frontal and parietal areas in spatial processing.

Regardless of lateralization, all findings show support for a “hard-wired” biological system that is developed specifically for processing spatial information. Evidence for such a system leaves no doubt that a SpA latent construct is valid. In addition to theory driven by cognitive science, this fact provides a rational foundation upon which human performance/ability investigations can be based. However, the

question of how to *properly quantify* SpA as a cognitive construct has been a point of interest for many psychometricians and is a primary focus of this research.

High-Level Definitions of SpA Constructs

SpA is an underrepresented component of the human cognitive repertoire in terms of application of psychometric outcomes relative to other factors such as quantitative and verbal abilities.

Nonetheless, scientists have recognized SpA constructs as an empirically central aspect of human intelligence as early as the beginning of the 20th century (Binet & Simon, 1916). Thurstone (1938) proposed that human intelligence is constituted of seven independent factors that he referred to as “primary abilities,” including word fluency, verbal comprehension, spatial visualization, number facility, associative memory, reasoning, and perceptual speed. More recently, Snow (1996) provided a simpler model, with intelligence being constituted by a complexity dimension and three content domains: quantitative/numerical, spatial/mechanical, and verbal/linguistic. In addressing the spatial component of intelligence specifically, Thurstone (1938) recognized that the SpA construct can be further divided into three sub-components: (a) object recognition from different angles, (b) imagining movement/displacement of constituents of a spatial configuration, and (c) determining spatial relationships with respect to one’s body. Subdivision of SpA has also been supported in the recent literature. Sternberg (2000) stated that SpA is not a unitary construct but possesses several subcomponents, each emphasizing different aspects of image generation, storage, retrieval and transformation. Subdivisions of the SpA construct have been well-developed by Carroll (1993) (and subsequently widely adopted by researchers in the field), who showed that the general spatial construct consists of a hierarchy of sub-constructs, in which the performance variance of a generalized construct is split among more specific SPA forms abbreviated here:

- General Spatial Visualization factor: The general spatial visualization factor (Gv or Vz) is at the hierarchy’s pinnacle due to relative complexity of processing; it can be measured and quantified by, among others, paper form board, paper folding, and mental rotation tests (e.g., Shepard-Metzler, 1971).
- Orientation (SO) factor: Research participants are asked to imagine how an array would appear from a different perspective and then to make a judgment based on the imagined perspective (e.g., an aerial orientation test).
- Rotation (SR) factor: This factor emerges if two or more simple, highly speeded mental rotation tasks are included in a test battery (e.g., flags, embedded figures, mental rotation; Shepard-Metzler, 1971 also fits here).

Sternberg (2000) indicated that Vz tests appear to be primarily measures of general intelligence, are secondarily measures of task-specific functions, and thirdly an “undefined” construct that covaries uniquely with other Vz tests. He went on to indicate

that SO factors can be distinguished from Vz factors, and that the SR factor emerges if two or more simple, speeded tasks are included in a test battery.

An Ecological Explanation for SpA Constructs

Spatial visualization and mental rotation cognitive functions serve as representations of one's own body in space, as it is understood in terms of one's positional relationship to environmental conditions (such as gravity), or to the location of other objects in space. The natural condition of orienting one's self with regard to the position of environmental objects has been explicated by Amorim, Isableu and Jarraja (2006).

They indicated that "spatial embodiment," or giving human characteristics to non-human geometric shapes, helps to improve performance in mental rotation tasks. They attributed this improvement in performance to the general idea that human spatial cognition is action-oriented, and is necessarily a function of taking into account the human body and its spatial and motor representations. Such a conception generally provides for the validity of investigations that consider mental rotation and spatial orientation-types of constructs. From a face-validity perspective, tests that require participants to perform tasks, which appear to require the same resources as real-life spatial activities, simply "fit" the action orientation of SpA. For construct validity, paper and pencil or computer-based (virtual) analogs of real-life conditions that demand those spatial resources as are employed in similar real-life conditions, can be shown to "fit" quantitatively. Perhaps it is this embodiment effect that provides spatial measures with generally strong face validity and exceptionally strong ecological validity as long as the testing task is an accurate reflection of relevant real-world conditions.

Gender Differences

Gender differences are an important aspect of SpA. A majority of modern SpA investigations have gender differences at their root. Historically, findings have shown performance advantages in males with regard to SpA metrics. An example is a recent study (Geary & DeSoto, 2001) which showed significant differences in performance between males and females, with males being "over-represented" at the high end of the performance scale in a battery of SpA tests, and females being "over-represented" at the lower end of the performance scale.

Results also showed that such differences were independent of culture (comparing participants from U.S. and Chinese populations), the implication being that genetic-evolutionary processes lie at the heart of SpA gender differences (Geary & DeSoto, 2001). An example of "spin-off" gender difference research is demonstrated in the work of Kass, Ahlers, and Dugger (1998), who showed that gender differences in visual SpA (estimation of orientation angle of a ship viewed through a submarine periscope simulator) could be reduced via training and feedback. Another example of such research is the work of Bowers, Milham, and Price (1998) who compared performance of males and females on different SpA tests and other tasks in an effort to infer differences in brain

lateralization by gender. Results indicated no systematic SpA differences, and the authors implied that there are no differences in lateralization processes. A further implication of this study has been that the SpA construct has not historically been measured correctly. Bowers et al. indicated that, given the multi-faceted nature of SpA, “inconsistencies” in the literature with regard to gender differences have been found because researchers normally use a single measure of SpA without identifying the SpA subtype it is intended to measure. Such a critique of approaches to measuring SpA is in line with Sternberg’s (2000) admonition to accurately measure a construct, but within the context of Bowers et al.’s (1998) work, the admonition also includes identifying the unique variance of SpA subtypes.

Given the two examples presented, gender differences are central to most discussions that involve SpA psychometrics. Research that involves the investigation of gender differences can have implications in the way that SpA constructs are conceived and in the way that they are subsequently measured.

The ultimate outcome of the debate can have repercussions within the realm of validity. For example, Bowers, Milham and Price (1998) investigated the degree to which lateralization of spatial processing, and in effect, gender differences, are artifacts of test selection. Their findings suggest that, prior to administration, tests of spatial ability should be screened for gender-related sensitivity given that completion of spatial ability tasks makes demands on varying levels of skills across measures. This outcome is explicit in Bowers et al.’s (1998) study, because the question of whether a specific measure of SpA, the Guilford-Zimmerman Spatial Orientation test, is an effective SpA measure was addressed in the study. The study confirmed the efficacy of the test.

Quantification of SpA

A consequence of not including SpA in current approaches to measuring human performance, according to Shea, Lubinski, and Benbow (2001), is that modern talent search procedures currently miss approximately 50% of the top 1% in three-dimensional spatial visualization. This finding has implications in various vocational domains in which SpA plays a role, such as surgery, engineering, mathematics, and flying, especially if a goal in such professional communities is to obviate potentially outstanding performers.

Spatial abilities have been measured with performance tests, paper-and-pencil tests, verbal tests, and film or dynamic computer-based tests (Lohman, 1993). Performance tests are among the earliest measures of SpA and have included form board, block manipulation, and paper-folding tasks, for example (Binet & Simon, 1916). According to Lohman (1993), individual differences on most spatial tasks appear to be well accounted for by performance on factors defined by paper-and-pencil tests. This is reflected by the fact that most spatial tests are administered via paper-and-pencil format. An example of the paper and pencil approach to measuring SpA is provided by the ASTB’s Spatial Apperception Test (SAT). Figure 1 provides an example SAT item. In the SAT, participants are asked to determine the orientation of an airplane based on the

rendering of a pilot's view of the horizon. The SAT is very similar to the historically more widely administered Guilford-Zimmerman (1948) Spatial Orientation test.

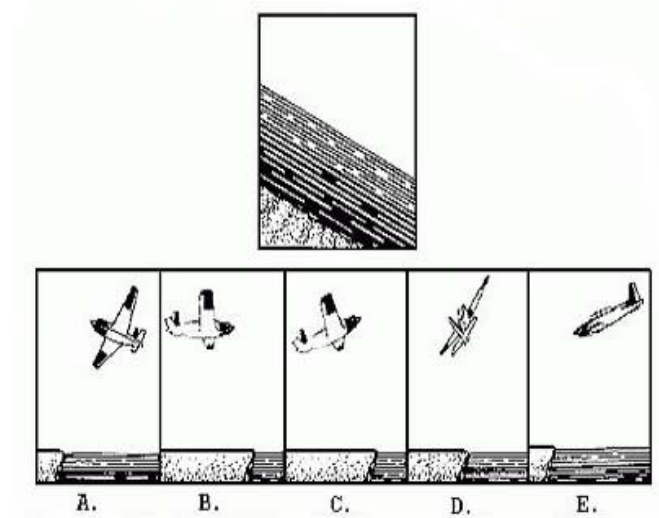


Figure 1. Sample Spatial Apperception item.

Verbal tests measure SpA by requiring participants to listen to a problem in which a mental model must be created, and to verbally respond. Although verbal tests of SpA are not often used, they have shown high correlations with other spatial tests and various criterion measures (Ackerman & Kanfer, 1993; Guilford & Lacey, 1947). Verbal measures of SpA often challenge the test taker to solve spatial problems posed in a scenario-based format. For example, Ackerman and Kanfer (1993) provided an example of the Verbal Test of SpA (VTSA).

In this test, participants are asked to close their eyes and imagine items described verbally (textually), and they are asked a multiple-choice question. The following is an excerpt from Ackerman and Kanfer's work in which the VTSA was administered:

It is morning and you are facing east looking at the sunrise. You walk forward for 100 yards, turn left, and after walking another 50 yards you turn about (i.e., turn 180 degrees). In what direction are you now facing?
a) North b) South c) East d) West (Ackerman & Kanfer, 1993, p. 431).

Some measures of SpA are more dynamic and attempt to tap into one's ability to adapt to object trajectories and/or arrival times, and often include a psychomotor performance component. An example of such a test is provided in Figure 2, showing a screen shot of a 2-dimensional compensatory psychomotor-tracking task. In this task, performance is measured as the participant's ability to place and maintain the cursor in the center of the screen, using a joystick control. Participants in this task must continuously adapt to random movements in the cursor however, requiring compensation for changing movements in the cursor that are beyond the participant's control. Such measures, by default, include an element of psychomotor ability given that responses cannot be made without behavioral input of the participant. However, it must

be noted that such measures are used to select personnel who possess highly effective combined spatial and psychomotor abilities for very specific tasks, such as military aircraft piloting.

In the early, through mid-, 20th century, the use of performance based/SpA tests for military pilot selection proved to be a valid effort in the prediction of primary flight training in the U.S. Navy (Fleishman, 1956).

Also, performance based/SpA tests were shown to be good predictors of success in training courses for aircrew positions (Guilford & Lacey, 1947). However, these tests were comprised of full-sized cockpit simulator-like apparatus, such as a stick-and-rudder control device or the Ruggles Orientator (Figure 3). As pilot selection testing moved away from military bases and onto college campuses, administering such tests became unfeasible due to the cumbersome nature of the devices and the difficulty of ensuring common calibration among apparatus.

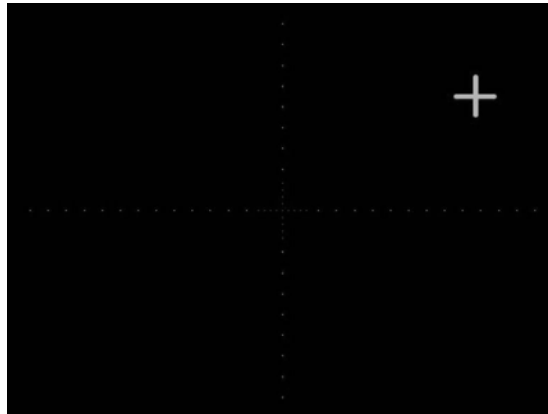


Figure 2. 2-D Compensatory Tracking screen shot.

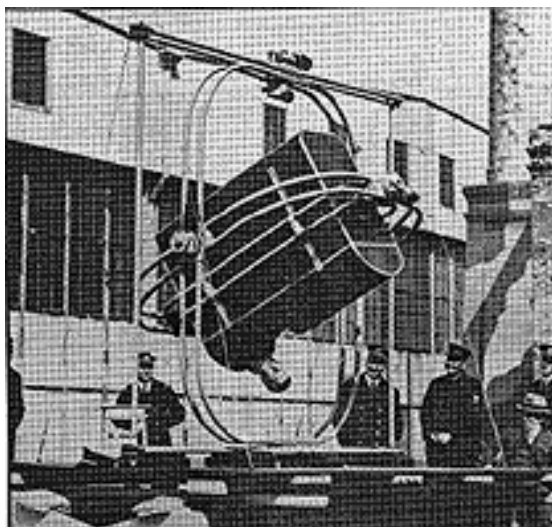


Figure 3. Ruggles Orientator.

Renewed interest in performance-based tests has accompanied advances in computer technology, as tests can now be administered remotely and results can be sent to military administrators via secure network connections. In addition, computer-based tests offer the opportunity to gather both error and latency scores, which can then be combined to predict criterion performances with greater precision than from either measure considered separately (Ackerman & Lohman, 1990). Research has been conducted demonstrating that computerized performance-based tests add a component of prediction to outcome based on ecological validity (e.g., Ackerman, 2001). Ackerman's research supports early work done in the military aviation selection arena. For example, Griffin (1987) showed that performance on computer-based dichotic listening and psychomotor tasks were significantly correlated with U.S. Navy primary flight training grades.

Carretta (1989) showed that computer-based tracking related more strongly with a pass/fail criterion in U.S. Air Force flight training than any other test that was used in a battery of tests used for pilot selection. Delaney's (1990) follow-on study to Griffin's work, but using a much larger sample, showed that computer-based dichotic listening and psychomotor tasks were significantly related to U.S. Navy primary flight training grades. These studies have provided evidence that computerized performance based measures have great utility in accounting for unexplained variance in training outcomes. The current study continued along that line, but with a focus in the specific area of SpA, with a particular interest in mental rotation.

Mental Rotation

For the purposes of this research, mental rotation was operationally defined as *the ability to rotate mental representations of two-dimensional and three-dimensional objects*. It has been established that a general SpA construct is constituted of sub-factors, such as those found in Thurstone's divisions of SpA, although the first empirical work that focused specifically on mental rotation, as a relatively orthogonal psychometric construct was not conceptualized until the early 1970s (Shepard & Metzler, 1971). An example of the experimental stimulus is provided in Figure 4. Results of their study (summarized in Figure 5) showed that recognition and psychomotor response latency of two-dimensional renderings of three-dimensional shapes were a positive linear function of the angular difference in the portrayed orientations and that this latency did not differ depending on the number of dimensions in which rotations occurred. That is, one-dimensional (1-D) rotation latency is not less or greater than rotation in three dimensions (3-D).

Shepard and Metzler's (1971) results also showed that the slope of the obtained functions represented an average rate at which the stimuli were mentally rotated of approximately 60° per second, including the time it took to execute psychomotor responses.

It is theorized that specific events occur during mental rotation. According to Johnson (1990), mental rotation can be separated into the following stages:

- Creation of a mental image of an object
- Rotation of an object mentally until a comparison can be made
- Make the comparison
- Decide if the objects are the same or not
- Report the decision

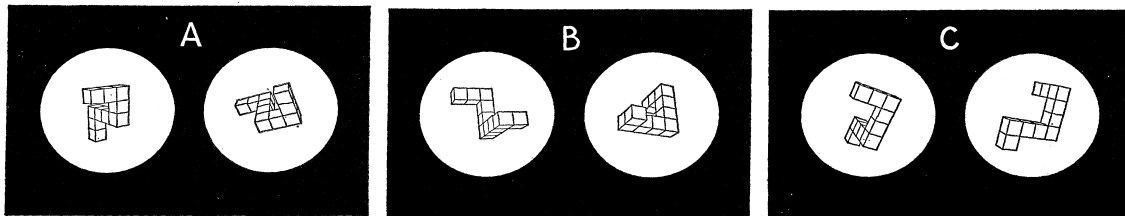


Figure 4. Mental rotation stimuli used by Shepard and Metzler (1971). Examples of pairs of perspective line drawings presented to the subjects. (A) A “same” pair, which differs by an 80° rotation in the picture plane; (B) a “same” pair, which differs by an 80° rotation in depth; and (C) a “different” pair, which cannot be brought into congruence by any rotation.

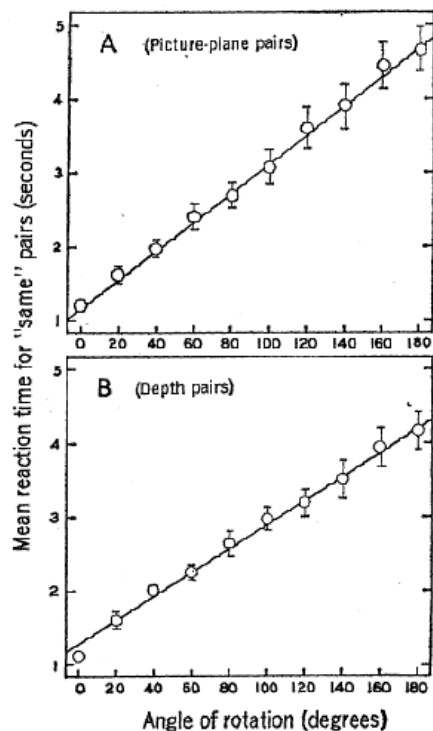


Figure 5. Chronometric results of Shepard-Metzler's (1971) experimental protocol

Psychomotor Ability Operational Definition

Psychomotor ability is defined as that which reflects the “*capabilities of the motor system to plan, coordinate and execute movements*” (Ghez & Krakauer, 2000, p. 653). Guilford (1958) examined psychomotor ability and defined numerous aspects of psychomotor functioning and indicated that functioning could be classified according to the following kinds of abilities: strength, impulsion, speed, precision, coordination, and flexibility. Of the aspects of functioning presented, those most pertinent to the current study include precision (of directed movements) and coordination (of different limb movements).

Associated Neurological Processes

Motor processing is the reverse of the sequence in the sensory system (Saper, Iverson, & Frackowiak, 2000). Motor planning commences with a “general outline” of intended behavior and is transformed into actual responses in motor pathways. Patterns of frontal neurons firing constitute the source of individual and complex motor actions. The motor pathways leaving the cerebral cortex have their origin in the primary motor cortex of the precentral gyrus. Neurons in the premotor cortex are associated with planning of movements and receive inputs from motor centers in the thalamus, primary somatosensory cortex and the prefrontal association cortex. Although the premotor area sends projections to other areas of the brain, of great import are projections to the primary motor cortex (Saper et al., 2000) and cerebellum. Essentially, the primary motor cortex sends projections down the spinal cord to synapse with motor neurons that connect to muscles, with the cerebellum being central in motor control.

Generally, psychomotor movements are controlled by two kinds of systems: feedback control and feedforward control (Ghez & Krakauer, 2000). In feedback control, signals from sensory organs are compared to a reference signal by a “comparator.” Error, if any, is corrected by a change in output by a controlling mechanism. Feedforward control is determined by information acquired before feedback sensors are activated. Accuracy in movement using feedforward control requires prior knowledge of stimulus characteristics so that appropriate movements can take place. Demands were placed on both of these systems by the psychomotor tasks that were used in the current study.

Application of Psychomotor Data

An example of how psychomotor data can be applied in personnel selection is provided by the work of Johnston and Catano (2002). They examined the predictive and incremental validity of three psychomotor ability measures: manual dexterity, finger dexterity, and motor coordination. The researchers combined these measures with cognitive assessments and found that the three psychomotor measures increased validity for selected technical and mechanical occupations. Their findings suggest that PBMs such as psychomotor tasks can improve predictions beyond the sole use of cognitive measures when job analyses reveal that a given assessment is relevant.

Although the aim of this particular study is not to demonstrate such ecological validity for the tasks that are being examined, this work should pave the way for follow-on work that explores relationships between the experimental tasks that are being examined here and performance in U.S. Navy flight training. Additionally, examination of performance on the experimental tasks should not be limited to flight training, but other applications (e.g., medical applications) should also be explored.

Computer-Based Medium

Spatial constructs can be measured digitally. Pelligrino, Hunt, Abate, and Farr. (1987) demonstrated a computer-based battery that assessed SpA factors and found that an integrated software package constituted a valid performance measurement system for the identified constructs.

When the measure of any construct has been administered using a paper-and-pencil medium, it is necessary to ensure that the validity of the measure is not lost when a change in administration medium has taken place. Larson (1996) showed that paper-and-pencil SpA tests can measure the same construct as computer-based tests.

Current technology enables test administrators to present dynamic visual displays to test takers and to compile large data set that have been created by users quickly and accurately. Computer-based testing media permit the presentation and collection of both static and dynamic data, making the inclusion of performance based measures possible at sites that are remote to central testing organizations. These same systems can reliably administer required tests with minimal calibration across different testing sites. For example, at selected administration sites the U.S. Navy is offering a web- and computer-based version of the ASTB, referred to as the Automated Pilot Exam (APEX). This version of the ASTB allows for administration in distributed areas, with a centrally-located secure server and scoring center (Naval Operational Medicine Institute, 2004). Notwithstanding the reduced time in test administration, scoring, and reporting, platforms such as APEX present an ideal means by which performance-based tests can be administered.

Summary

Spatial and psychomotor ability are constructs of interest in Naval Aviation personnel selection given that such abilities are critical for the safe operation of aircraft and for accurate navigation. Additionally, such latent constructs could prove to be the source of untapped explained variability with regard to human performance measurement, which if tapped, could yield significant savings in terms of training dollars.

The purpose of the current work was to determine if the selected experimental tasks (i.e., BRT and 2-D Compensatory Tracking) demonstrate an improvement in explaining performance variance as they relate to validated tasks via examination of potential incremental validity of experimental tasks. More specifically, the BRT was investigated for utility as a measure of both mental rotation and psychomotor ability,

reflected by association with Compensatory Tracking (a measure of psychomotor ability), and the Mental Rotation Test (a measure of mental rotation). Results addressed reliability coefficients and correlations among experimental spatial/psychomotor ability measures (BRT and Compensatory Tracking) and existing, validated measures (Vandenberg Mental Rotation, SAT, and Manikin; Vandenberg & Kuse, 1978). Finally, descriptive performance models were created for both the BRT and the 2-D tracking task.

CHAPTER 3: METHOD

Participants

Students from the University of Central Florida (UCF) were solicited to volunteer for this study via the UCF Psychology Department and in accordance with Federal law and with the requirements of the University of Central Florida's (UCF) Institutional Review Board (IRB). All participants were briefed on the protocol, informed of the voluntary nature of participation and asked to sign an Informed Consent form. All participants were assigned a randomly generated participant number to protect privacy. No personal identifiers were used on materials that contain participant data. All data were maintained in a secure location. Instructors were asked to provide extra credit in Psychology courses for incentive for students to participate. Cohen (1992) indicates that $N = 42$ is sufficient for power = .80 and $\alpha = .05$ to detect large effects of 5 variables using multiple regression/correlation (MRC) analyses. To meet and exceed power requirements, more than 70 students' participation was accessed.

Research Design

A central focus of this work was to determine if the BRT produces an increased explanatory account in performance variance, indicated by comparison of scores on the BRT (percent correct), Compensatory Tracking (mean deviation from target), Spatial Apperception (percent correct), Mental Rotation (percent correct), and Manikin (percent correct) tasks. Each participant is asked to complete a battery of tests consisting of a pre-simulation questionnaire, the BRT, 2-D Tracking, SAT, MRT and Manikin. Gender differences in performance will be examined to determine if performance on spatial ability concords with gender differences that pervade the literature.

Gender differences were examined to demonstrate the validity of tests in the sense that findings in this study are similar to those found in other studies. Second, a correlational analysis was conducted to determine the independence or relatedness of task performance to infer corresponding latent construct behavior. This analysis provided the basis for tests of incremental validity. That is, correlational data from task performance on primarily the BRT, Mental Rotation, Spatial Apperception and Compensatory Tracking tasks showed whether mental rotation and psychomotor ability can all be measured with the BRT. If the administered tests that were presumed to measure different constructs (psychomotor ability vs. mental rotation) expressed no relationship, independence of constructs was indicated. Normative data and trends were also sought with regard to performance on the novel experimental tasks. This aspect of the investigation was also intended to explicate, at a fundamental level, the reliability of BRT and Tracking measures.

Table 1

Constructs and Previously Validated Tasks

Construct	Validated Task
Visualization	Vandenburg Mental Rotation
Orientation	Spatial Apperception
Rotation	Vandenburg/Manikin
Psychomotor	2-D Compensatory Tracking

Materials

Descriptions of the tasks to be used include:

Pre-Simulation Questionnaire (adapted from Qu, 2003) (Appendix B)

This survey solicited primarily demographic information from the participant as well as information about actual flight experience and experience using Microsoft Flight Simulator. Approximate administration time: 5 minutes.

Spatial Apperception (SAT)

Paper-and pencil administered test in which participants select an outside static view of ground-aircraft relation based on static presentation of horizon as seen from the cockpit. Correct identification of targets requires spatial relations and orienting capabilities. Performance is measured in terms of the total number of correct items. Participants must complete 25 items in 10 minutes.

Block Rotation Task (BRT) (Figure 6)

The BRT is an automated variant of the Shepard and Metzler (1971) mental rotation task. On the left side of the computer screen, the 3-D BRT presents a static 3-D figure constructed from blocks, referred to as the “goal figure.” An identical comparison figure is presented on the right side of the screen, but it is in a different orientation in terms of its rotation along any one of the x, y, and z axes individually or in combination. The objective is to use a joystick and throttle to rotate the comparison figure into exact alignment with the goal figure. After administering a practice session, the computer allows one minute to solve each of 24 problems and records the time required (up to one minute) to solve each problem. The computer also records error (deviation of the comparison figure from the target figure) in the subject’s solution in each of the three axes. Error is measured on a 100-point scale, where 1.00 is the greatest distance between the rotational axis of the movable figure and that of the target figure and 0.00 is a perfect match between the two. Directionality of separation between axes is indicated by positive and negative values.

If the participant does not solve the problem in the allotted time, the computer automatically steps to the next problem and scores the missed problem as an error. Otherwise, the subject can elect to move on to the next problem if he/she is satisfied with the solution. Performance is measured in terms of the total number of correct items as

well as other descriptive data (e.g., combined spatial and psychomotor human performance modeling) that will be explored for the first time ever here in this study. Sampling rate is 500 ms. Approximate administration time: 25 minutes.

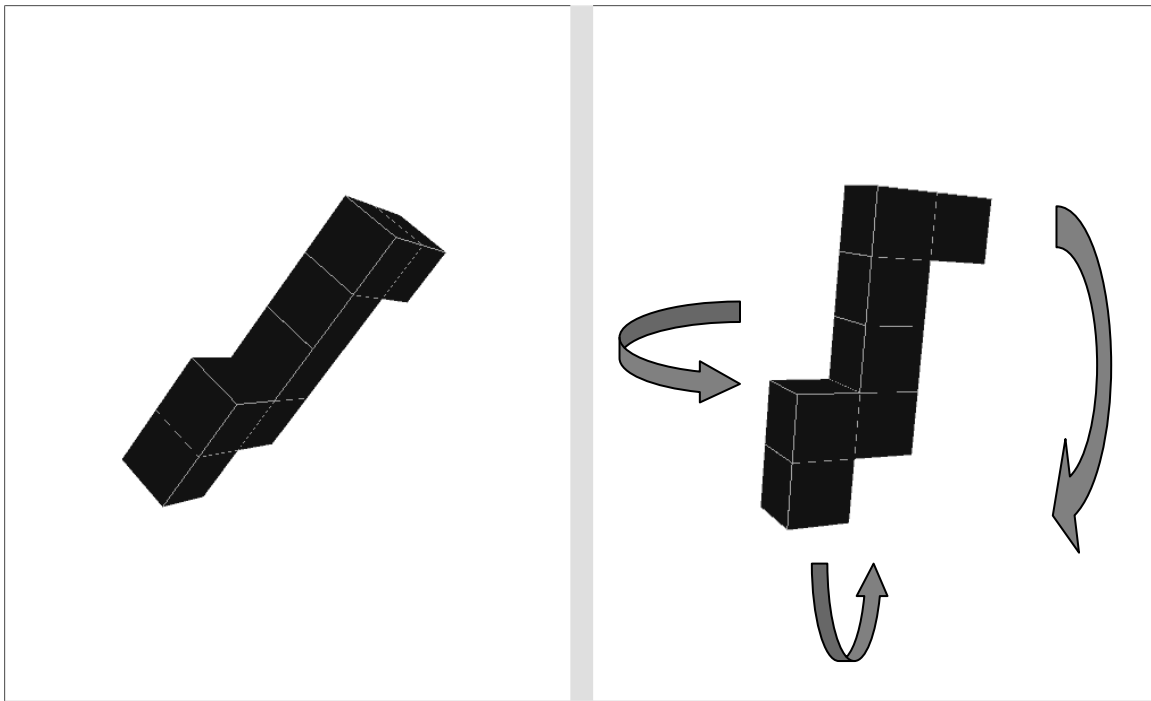


Figure 6. Sample Block Rotation Task item

Vandenberg Mental Rotation Test (MRT)

This paper-and pencil administered test's 20 items is organized into five sets of four. Items consist of a criterion figure, two correct alternatives and two incorrect distracters. Correct alternatives are identical to the target shape. In half of the items, distracters are rotated mirror images of the criterion while in the other half, distracters are rotated images of criteria in other items. Items are scored correctly only if both correct alternatives are identified and are considered incorrect if both correct alternatives are not identified. Performance is measured in terms of the total number of correct items. Approximate administration time: 10 minutes.

Manikin Test (Lane & Kennedy, 1990)

In this computer-administered test, participants must determine which hand, right or left, is holding the object that matches the object on which the manikin is standing. The manikin may be positioned standing upright facing either toward or away from the participant, or upside down, also facing toward or away from the subject. The manikin's position can be distinguished by characteristics such as facial features and clothing. Responses are made by pressing an arrow key, where the arrow pointing left is indicative of the object being held in the left hand and the right arrow represents the object being held in the right hand. Performance is measured in terms of the total number of correct items. Approximate administration time: 1.5 minutes.

2-D Compensatory Tracking (Tracking)

This is a computer-administered task in which participants are asked to keep a set of randomly drifting crosshairs in the marked center of a computer display screen. Scoring consists of a 100-point scale using a traditional Cartesian coordinate system (that includes negative values), where 1.00 is the greatest distance between the crosshairs and the center of the screen and 0.00 is the least distance between the two. Sampling rate is 100 ms. Approximate administration time: 5 minutes.

Procedure

Participants were solicited via the required UCF on-campus recruitment procedure (Sona Systems), were briefed and asked to read and sign the consent form, and were then administered the pre-simulation survey. Experimental tasks were administered to participants via laptop computer (Dell Latitude D620). All participants were briefed on the protocol and were asked to read and sign the UCF-approved Informed Consent Form.

The BRT, Spatial Apperception Test (SAT), Vandenberg Mental Rotation Test (MRT), Manikin (M) test and 2-D compensatory tracking task (Tracking) were administered in a generally counterbalanced fashion. However, due to constraints inherent to the novel software, BRT and Tracking were administered in the same order. After being assigned a randomly-generated participant number to protect privacy, participants were administered the pre-simulation questionnaire and the spatial abilities battery (~60 minute administration time total). Upon completion of the battery of tasks, all participants were debriefed and thanked for their effort.

CHAPTER 4: RESULTS

Data Screening Descriptive Statistics

Data analysis consisted primarily of descriptive statistics, multiple regression/correlation and ANOVA. Other tests were conducted on an as-needed basis. Standard level of significance for statistical tests was set at $\alpha = .05$. Statistical tests were run using SPSS 10.0. Power analyses were conducted using G-Power 2.0 (Buchner, Erdfelder, & Faul, 2001).

Basic demographic and performance descriptive statistics are presented in Table 2. Table 3 shows handedness distributions across gender, where data in both samples approximate the population norm, where it is generally understood that 90% of the population is right-hand dominant.

Data from 73 pseudo-randomly (solicitation was completed via the UCF Psychology Department) solicited participants were collected, consisting of $n = 52$ (71%) females and $n = 21$ (29%) males from UCF's undergraduate student population. The male-female proportion disparity may be explained by the fact that there are generally more females than males enrolled in universities across the U.S. Further, the disproportion is perhaps exacerbated by the fact that there are generally more females than males enrolled in Psychology courses.

Most participants completed all tests. However, some participants did not complete all assessments due to apparently random errors related to understanding or following instructions or due to apparently random errors related to computer operation. However, as reflected by descriptive statistics that accompany each particular analysis in this study, such discrepancies are minimal and do not appear to pose a threat to the validity of these results.

BRT Data Screening and Cleaning

If one wished to examine BRT performance only in terms of the total number of correct items, one would be at a loss for the richness of data that this assessment can provide. The total number of correct items on any validated task can provide valuable information for anyone who wishes to have "quick and dirty" human performance information, and the BRT may well provide that. However, given the complexity of the interaction of *combining* human performance abilities (and in the case of the BRT this appears to be spatial ability and psychomotor performance) a tool that measures such interactions is bound, perhaps even required, to produce data that approach the complexity of the behavior itself. Such is the case with the BRT. The aims of this study required the use of the total number of correct items for correlational analysis as well the use of collective performance data to produce a "normative" human performance model for 3-D processing and associated psychomotor behavior. Total number of correct BRT responses was uncomplicated to derive. However, the BRT software produces an

enormous amount of data per participant per testing session. For example, one item consists of approximately 60 data sets if a participant is engaged in the item for the entire 30 seconds. Multiplied by 28 (the total number of items per test session), this provides the investigator with almost 1700 sets of raw performance data per participant per testing session if each participant takes the full 30 seconds to complete each item.

Graphs representing the completion of a single BRT item follow, illustrating the process used to screen BRT data for normative analysis. Graph lines represent error between the BRT axes (3 dimensions) as well as an additional line representing the mean.

Table 2

General Descriptive Statistics

Variable	N	Min.-Max.	Mean (M)	Standard Deviation (SD)
Females	52	---	---	---
Males	21	---	---	---
Age	73	18-32	19.44	2.30
Manikin Total				
Correct	69	13-67	40.53	11.01
Tracking Mean				
Deviation	73	.17-.81	.39	0.13
Tracking Standard				
Deviation	73	.14-.38	.26	0.58
BRT Total				
Correct	73	0-19	3.48	4.27
SAT Total				
Correct	73	0-22	10.23	4.84
MRT Total				
Correct	72	0-20	9.04	5.47

Table 3

Participant Gender-Handedness Demographics

<u>Male</u>		<u>Female</u>	
Left-Handed	Right-Handed	Left-Handed	Right-Handed
2 (10%)	19 (90%)	7 (13%)	45 (87%)

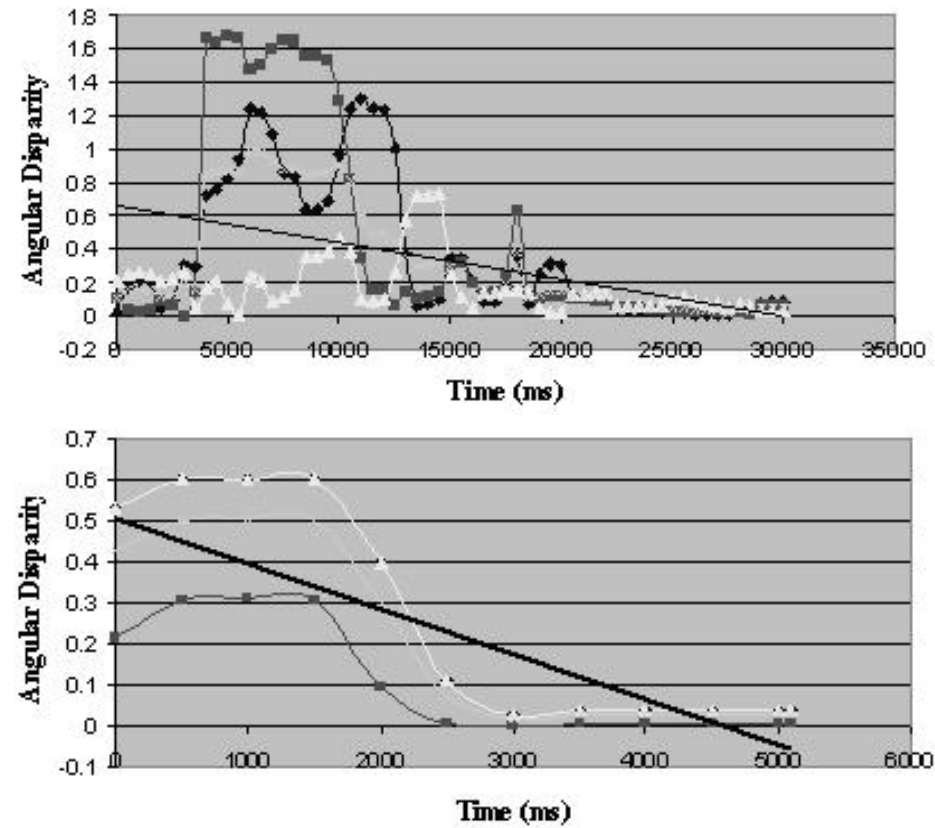


Figure 7. Ideal Performance. The graph on the top characterizes an ideal and most frequently encountered form of BRT error reduction, with an initial perturbation and a gradual reduction of error across all three axes. On the bottom, a characteristic reduction is shown but completed in relatively short order (only 5 seconds to completion).

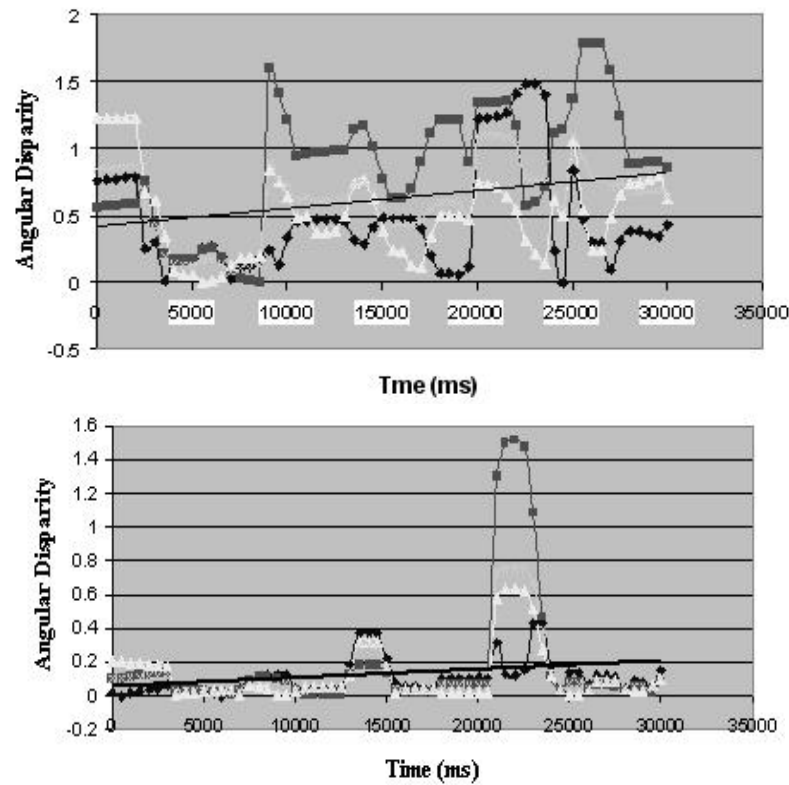


Figure 8. Less Than Ideal, Corrective. Where Figure 7 shows the most common representation of error reduction, these graphs demonstrate the variability with which BRT items can be corrected. The characteristic initial perturbation in Figure 7 is absent, but there is a gradual error reduction as time advances. Evidenced by the normative results, these kinds of data dilute what could be a more linear relationship between time and error reduction. However, given that behaviors resulting in “atypical” data such as these resulted in successful error reduction, the data were entered into the BRT normative performance model.

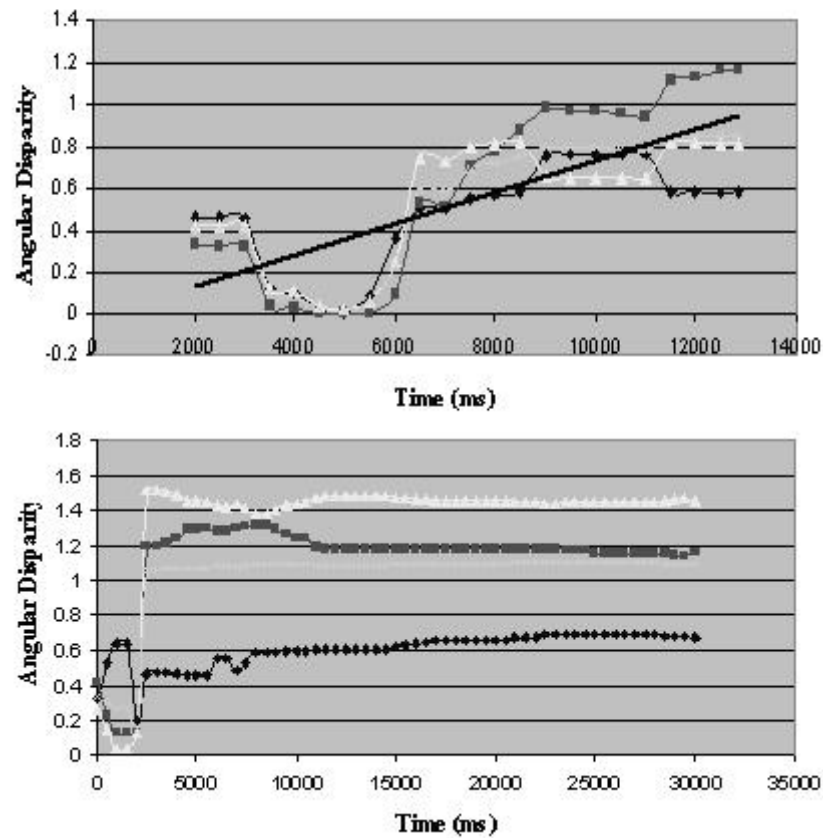


Figure 9. Non-Corrective. Juxtaposed to the examples of error reduction provided in Figures 7 and 8, these graphs show the absence of error reduction altogether. The graph on the top shows that the participant manipulated computer peripherals but was unsuccessful in any error reduction prior to cessation of problem-solving. The graph on the bottom shows no manipulation of computer peripherals indicating the complete absence of an attempt at problem-solving. Data such as these were not included in the normative analysis given that the goal of the normative analysis was to model human performance that reduces error between block stimuli.

Assessment Reliability

Any inference based upon measures made by an unreliable tool is invalid. To address this issue, reliability assessments were conducted on the performance assessments that were used in this study where possible. If a reliability analysis was not conducted for a particular assessment, the raw data were such that it was not possible to properly conduct the analysis. In cases where a reliability analysis was not conducted, reliability findings are cited or are inferred from the raw data to the extent possible.

Reliability estimates for the SAT and MRT in this study were made by transforming correct and incorrect item responses into “dummy variables” (1 = correct and 0 = incorrect). Cronbach’s alpha and split-half analyses were conducted for both assessments with results displayed in Table 4. Findings indicated marginal to adequate reliability coefficients for both the SAT and MRT. Lane and Kennedy (1990) indicate that an unreliable test is one with inter-trial correlations of about .70, which may involve too much error measurement to be useful in repeated measures designs. It is notable that the SAT’s reliability coefficients were not higher. Assessments in this study were administered to participants only once, so any notion of measurement stability could not be addressed. However, a repeated measures approach was emulated by conducting the specified reliability estimate procedures.

A reliability estimate for the Manikin test was not conducted because the software does not make available data for each item response. This is also the case for the BRT in terms of providing data for correct vs. incorrect data, but the richness of the raw data from the BRT can have a compensatory effect.

Although reliability data were not available for the Manikin task, Lane and Kennedy (1990) provide some reliability information. They indicate that the Manikin task possesses a “reliability efficiency” estimate of .91, where reliability efficiency is defined as the (normalized) largest reliability likely to be encountered in practical applications.

Reliability data for the Tracking and BRT assessments were not conducted and perhaps will be conducted in a future study. However, the reliability of these tests may be inferred from their intercorrelations with validated tasks and via inspection of the descriptive models that will be provided when Hypothesis III is addressed.

Approach and Analysis Gender Differences

Tables 5 and 6 show significant gender differences in all measurement domains with the exception of Tracking Standard Deviation, with males having demonstrated superior performance. This finding concords with most extant literature that examines gender differences with regard to spatial processing and serves as one aspect of the set of indicators in this study that the measures used here were valid SpA assessments. However, a post-hoc power analysis indicated that this test possesses a statistical power

of about .56. In order to achieve the desired traditional power of .80, approximately 130 participants were needed.

Results as they relate to the hypotheses of this study are provided.

Hypothesis I

In line with the expectation that the BRT is a measure of both psychomotor ability and mental rotation, performance on both the psychomotor task (Compensatory Tracking) and the mental rotation tasks (Spatial Apperception, Mental Rotation, and Manikin) were expected to positively correlate with the BRT. To demonstrate that the BRT is a measure of two independent constructs, the psychomotor task (Tracking) was expected to be uncorrelated with any SpA measure. Table 7 shows the correlation matrix supporting the hypothesis that the BRT measures both spatial and psychomotor abilities. However, given that the psychomotor task correlated with all SpA measures, independence of psychomotor (via 2D Tracking) and SpA constructs was not demonstrated using the performance assessments of this study. Power analysis for an r effect size of .30 with a sample size of 73 indicated β at 0.84.

Table 4

Reliability Analysis for SAT and MRT

Assessment	N of Cases	N of Variables	M	Variance	SD	Alpha	Split-Half*
SAT	63	25	10.68	23.16	4.81	.79	.65
MRT	65	20	9.60	28.28	5.32	.88	.72

*Note: Between-forms correlation.

Table 5
Analysis of Variance Descriptive Data

<u>Assessment</u>	<u>Gender</u>	<u>N</u>	<u>M</u>	<u>SD</u>	<u>Minimum</u>	<u>Maximum</u>
Manikin Total Correct						
	Female	49	38.69	10.23	13.00	61.00
	Male	20	45.30	11.93	27.00	67.00
Tracking Mean Deviation						
	Female	52	0.41	0.13	0.20	0.81
	Male	21	0.33	0.12	0.17	0.63
Tracking Standard Deviation						
	Female	52	0.27	0.05	0.16	0.38
	Male	21	0.24	0.07	0.14	0.37
BRT Total Correct						
	Female	52	2.21	2.84	0.00	12.00
	Male	21	6.62	5.54	0.00	19.00
SAT Total Correct						
	Female	52	9.25	3.91	0.00	20.00
	Male	21	12.52	6.18	3.00	22.00
MRT Total Correct						
	Female	51	7.02	4.62	0.00	17.00
	Male	21	13.67	4.43	5.00	20.00

Table 6
Analysis of Variance for Gender Differences

<u>Source</u>	<u>Groupings</u>	<u>Sum of Squares</u>	<u>df</u>	<u>Mean Square</u>	<u>F</u>	<u>p</u>
Manikin Total Correct	Between	619.83	1	619.83	5.38	< .05
	Within	7726.60	67	115.32		
	Total	8346.43	68			
Tracking Mean Deviation	Between	0.09	1	0.09	5.91	< .05
	Within	1.07	71	0.02		
	Total	1.16	72			
Tracking Standard Deviation	Between	0.007	1	0.007	2.12	> .05
	Within	0.23	71	0.003		
	Total	0.24	72			
BRT Total Correct	Between	290.59	1	290.59	20.16	< .05
	Within	1023.63	71	14.42		
	Total	1314.22	72			
SAT Total Correct	Between	160.33	1	160.33	7.39	< .05
	Within	1540.99	71	21.70		
	Total	1701.32	72			
MRT Total Correct	Between	657.23	1	657.23	31.48	< .05
	Within	1461.65	70	20.88		
	Total	2118.88	71			

Table 7

Human Performance Correlation Matrix

Ability Measure	1	2	3	4	5	6	7
1. Manikin Total							
Correct	--	-.21*	-.11	.46*	.32*	.35*	-.08
2. Tracking Mean							
Deviation	--	--	.76*	-.31*	-.26*	-.24*	.11
3. Tracking Standard							
Deviation	--	--	--	-.18	-.17	-.17	.02
4. BRT Total							
Correct	--	--	--	--	.54*	.50*	.22*
5. SAT Total							
Correct	--	--	--	--	--	.47*	.03
6. MRT Total							
Correct	--	--	--	--	--	--	.02
7. BRT Mean							
Completion Time	--	--	--	--	--	--	--

*Note: $p < .05$, one-tailed.

Hypothesis II

Not unlike Hypothesis I, it was expected in Hypothesis II that performance on validated tasks would be predictive of performance on the BRT. The correlation matrix in Table 7 establishes that performance on validated tasks is related to that of BRT performance. In the matter of prediction, stepwise regression shows, in Table 8, the degree to which the 3 “best” models predict variance in BRT performance. With an effect size of $f^2 = 0.15$ and sample size of 73, power analysis showed that multiple regression indicated the following β values for (a) one predictor: 0.90, (b) two predictors: 0.83, and (c) three predictors: 0.78.

Table 8

Stepwise Regression Results

Model	R^2	R^2 Change	Change in p
1	.32	---	---
2	.40	.08	.005
3	.44	.04	.04

Notes: All models are significant predictors of BRT performance

Model 1 predictor: SAT Total Correct

Model 2 predictors: SAT Total Correct + Manikin Total Correct

Model 3 predictors: SAT Total Correct + Manikin Total Correct + MRT Total Correct

The regression results with regard to Hypothesis II show that the single best predictor of BRT performance is SAT performance. The model with the next highest predictive power is SAT performance combined with that of Manikin performance, providing an additional 8% of explained BRT performance variance. Finally, MRT added to the Model 2 provides an additional 4% explanation in BRT performance.

Gender differences with regard to predictive models were examined using stepwise regression. Results are shown in Tables 9 and 10.

Table 9

Female Stepwise Regression Results

Model	R^2
1	.30

Notes: Model is a significant predictor.

Model predictors: MRT Number Correct + Mean BRT Completion Time + Tracking RMSE + Manikin Number Correct + SAT Number Correct.

Table 10

Male Stepwise Regression Results

Model	R^2	R^2 Change	Change in p
1	.49	---	---
2	.59	.10	-.02

Notes: Models are significant predictors.

Model 1 predictor: SAT Number Correct.

Model 2 predictors: MRT Number Correct + Mean BRT Completion Time + Tracking RMSE + Manikin Number Correct + SAT Number Correct.

Comparison of models showed that a single model consisting of all tests used in the experiment predicts BRT performance for females and that two models, one consisting solely of SAT (spatial rotation) performance and the other consisting of all tests used in the experiment, predicts BRT performance in males. Analysis showed that the predictor models are stronger for males than for females ($R^2 = .49$ and $.59$ vs. $R^2 = .30$, respectively). It is notable that in prediction of male performance, Model #1, consisting only of SAT performance, was a relatively high predictor given that the entire set of tests used in the experiment, constituting Model #2, increased R^2 by only $.10$.

A hierarchical regression was performed to examine the potential confound effect that time to complete BRT items could have had on these results. The results of this analysis are presented in Table 11. Power analysis for the last regression procedure holds true for the current one.

Table 11

Hierarchical Regression Results

Model	R^2	R^2 Change	Change in p
1	.45	---	---
2	.49	.04	.04

Notes: Both models are significant predictors of BRT performance

Model 1 predictors: Tracking + MRT Total Correct + SAT Total Correct + Manikin Total Correct

Model 2 predictors: Tracking + MRT Total Correct + SAT Total Correct + Manikin Total Correct + Mean BRT Item Completion Time

Hypothesis III

It was expected that normative data for both the BRT and the Compensatory Tracking task would be derived from the data. With regard to the BRT, it was expected that a linear performance model could be derived from the BRT data similar to that of Shepard-Metzler (1971), where the behavior of solving a spatial-psychomotor problem can be chronometrically defined. The model in the current investigation was found to be significant in terms of error reduction with regard to time. The model constrained to linear form is shown in Figure 10.

Figure 11 shows raw Tracking performance data for all participants. The mean location along the x axis with regard to cursor location across all participants was -0.11 ($SD = 0.32$). The corresponding value for the y axis was -0.007 ($SD = 0.33$). These data indicate that participants tended to place the cursor slightly left and slightly below the target.

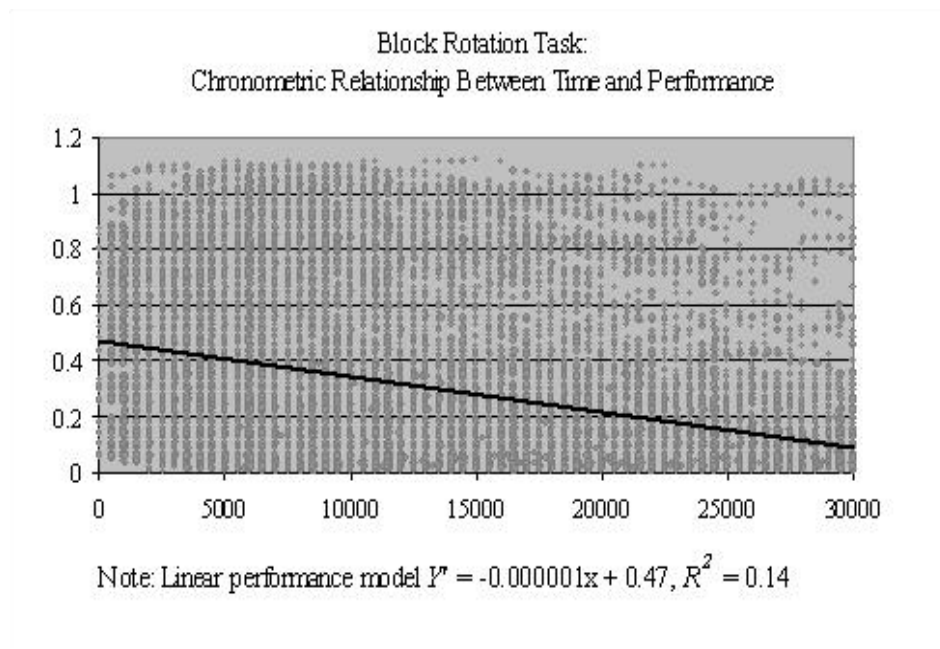


Figure 10. Chronometric relationship between time and error reduction on the BRT ($p < .05$). Data are shown for all participants and all test items.

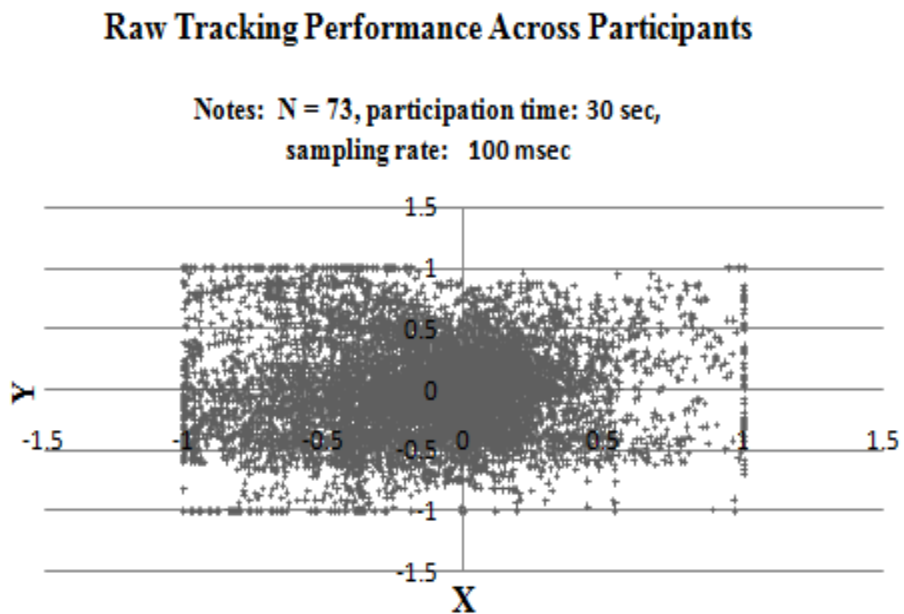


Figure 11. Tracking Performance Data. Tracking data across all participants are shown.

CHAPTER 5: DISCUSSION AND CONCLUSION

Initial Remarks

This study sought to identify possible BRT latent variables, to determine predictability of BRT performance, and to provide descriptive data for BRT and Tracking tasks. Those goals were achieved. However, before entering a discussion of results, it is necessary to point out a few items that will help to place the results in perspective. First, although a great effort was made to screen data for the different analyses that were presented, more can be done to produce a cleaner human performance model with regard to the BRT. There was so much variability in the manner in which successful BRT problem navigation was accomplished (as evidenced in Figure 7 and Figure 8) that valuable information would have been lost if models not conforming to the “ideal” expression had been not included in the analyses. However it is possible that the successful completion models that did not conform to the “ideal” form can show a number of possible alternatives to problem solving in the BRT context for future study. Further, different approaches to problem solving, expressed in “other than ideal” forms may have been a function of item difficulty, an area of interest for item analysis.

Assessments that had been previously validated (MRT, SAT and Manikin) have either shown suitable reliability coefficients in the current study or have shown acceptable reliability in the literature. The reliability of the BRT and Tracking tasks were not examined in this study with a formal test. This was not completed, in part, due to the unavailability of data with which to conduct such analyses. However, the model presented in Figure 10 is constituted of well over 100,000 individual data points (Sampling Rate (60/item) x N Items x N Participants), and Figure 11 is constituted of approximately 22,000 individual data points Sampling Rate (300/participant) x N Participants). Until suitable stability and reliability analyses can be conducted on these performance assessments, it is not unreasonable to presume an acceptable “working” reliability for the purposes of this investigation based on intercorrelations of these assessments with previously validated tasks and the appearance of the data that has been shown.

Hypothesis I

Hypothesis I was not fully supported by the results. Although strong correlations existed between the BRT and representatives from SpA and psychomotor domains, SpA and psychomotor domains were not demonstrated to be independent from one another, and did not truly represent the constructs intended. It is possible that Tracking task is not independent enough from spatial processing to be considered an observable psychomotor task. The Tracking task, by nature, requires participants to determine spatial relationships in order to maintain the movable cursor over the target cursor.

However, it is fitting that the BRT correlated with the SAT and MRT to a degree higher than that of the BRT’s correlation with Tracking and Manikin. The SAT and MRT are validated SpA measures whereas Tracking and Manikin are not SpA measures,

per se. Further, the strength of relation between the BRT and the other validated tasks showed that the BRT is *at least* a strong SpA measure.

Hypothesis II General Discussion

In hypothesis II it was presumed that validated tasks would predict BRT performance. Models were provided in which it was shown that the SAT spatial orientation task was the single best predictor of BRT performance. With Manikin included, an additional 10% of BRT performance variance can be explained; and when MRT performance is included in a model containing SAT, an additional 4% of BRT performance variance can be explained. This result is interesting given that Manikin is considered to be primarily a test of mental processing speed. To investigate the relationship between the time function and BRT performance, an hierarchical regression was performed using a variable consisting of mean BRT completion time. A model that included BRT item completion time only added 4% explanation in the variance of BRT performance, indicating that BRT item completion time was not a good predictor in relation to other tasks that were used in this study. However, mean BRT completion time was significantly correlated with BRT performance ($R = .22$), indicating that participants who took longer to complete an item generally outperformed those who did not.

Gender

The notion that females use different processing style to solve spatial problems was supported in this study given that males' and females' performance on the BRT could not be predicted using the same performance metrics. That is, female performance on the BRT was predicted by a variety of factors given that the predictive model of female performance included performance on all assessments that were used in the study. Conversely, a single best predictor was shown in the male data, where the SAT, a measure of spatial orientation, proved to be a better predictor than all the assessments combined. However, as indicated by Model #2 for males, all assessments were shown to be a significant predictor of BRT performance as well.

These results may reflect that different processing styles can impact assessment of spatial ability as it is currently being conducted in the U.S. Navy. The results of the different predictive models produced by males and females with regard to BRT performance showed that, at least in the instance of BRT performance, different approaches are used by the genders to solve spatial problems. Based on these results, the findings suggest that performance on the BRT can be explained by both simple and complex models whereas female performance can only be explained by a complex model. This finding supports the work of Bowers, Milham and Price (1998), who indicate that female spatial processing is more “distributed” than that of males.

Hypothesis III

The correlation between time and performance on the BRT model is significant. It shows that, on average, for every 0.47 msec there is a disparity reduction of 0.000010. This would be useful to know for any attempt at further validation, but based on data

screening, findings suggest that the true chronometric relationship is more likely curvilinear, with an initial perturbation that often causes more error than error reduction, and a subsequent, generally “exponential” error reduction. The initial perturbation in the data derived in this study was presumably a means for participants to derive a better understanding of how the block arrays were differentially configured. After obtaining this information, participants were presumably better equipped to reduce error between the block arrays, as suggested by the relatively rapid tapering off of disparity.

Data in the BRT chronometric model for the purposes of this study were constrained to a linear analysis. It is also possible, and likely, that the interaction between psychomotor and SpA processing is a non-linear function, exemplified, for example, by performance such as that depicted in Figure 7. However, a simple linear model is appropriate for the purpose of providing an initial glimpse of the BRT’s psychometric properties.

Tracking data showed that participants were generally adept at keeping the cursor in the general vicinity of the target, although the control reversal and variable movements assist in creating variability in the data.

Time and Performance

Notwithstanding the significant contribution of the “speed of processing” construct to performance in measures of general cognitive ability (Sternberg, 2000), it is necessary to describe the relationship between time, spatial processing and psychomotor performance and how they relate to the results of this study. Having roots in Gibson’s (1958) notion of ecological invariants in the visual flow field during a perception in action event is Lee’s (2006) theory of perceptuomotor control based on the ecological invariant, tau (τ). General τ theory posits that the organism acts a unitary entity in dynamic relations with the environment, highlighting interaction between the organism and its environment, and does not consider that organism as a complex mechanical device reducible into analyzable parts. Theory development has been driven via focus on relational or ecological invariants in engagements between organism and environment and is applied within the context of prospective guidance of movement, including “internal movements,” by means of the patterns of flow in sensory systems and activity patterns within the nervous system (Lee, 2006).

Central to the theory is the conception that movement is guided by “ τ -coupling motion-gaps.” Purposive movements of the body, and thoughts about such body movements, require guided closure of motion-gaps. A motion-gap is defined as the “changing gap between a current state and a goal state,” not unlike the condition imposed by the BRT of the current study. τ of a motion-gap is the first-order time-to-closure of the motion-gap or the current size of the motion-gap divided by rate of closure (Lee, 2006).

Such an approach to understanding the psychological phenomena that occur during task completion, like those imposed by the PBMs of the current study, concords

with a neuropsychological approach to understanding. Psychomotor movements being controlled by feedback control systems, for example, are no doubt influenced by a τ -type process, if not τ specifically. Sensory organs associated with feedback control systems permit comparison, at “higher levels,” of stimuli to a reference by a “comparator,” including the processing of τ -type data (i.e., an estimate of time-to contact) that could be used to modulate controlling mechanisms for error correction. Merchant and Georgopoulos (2006) provide a review and detailed descriptions of how τ plays a role in the neuropsychology of spatial and psychomotor processing. In relation to the current experimental context, τ or a similar measure may prove to be a useful means by which novel PBM performance data are transformed in future studies to compare BRT performance data with that of a similar ecological performance domain (e.g., flight training).

Study Limitations

It should be noted that participants in this study were university students. Although all Navy flight students are college graduates, they constitute a qualitatively unique subpopulation in comparison to university students. It is not known whether the findings here would mirror the results of the same study in which U.S. Navy Aviation students participated.

Given that gender differences have been expressed in the data here, it is possible that, without corrective or compensatory action (e.g., weighting), that the BRT could be perceived as discriminatory based on gender.

Another limitation of this study is that it will not provide a wide spectrum of evidence regarding application of other performance-based measures as good predictors of flight training performance. The focus of this research was in regard to SpA only and does not apply to other factors such as general psychomotor ability. Further, this study identified a specific aspect of SpA in a constrained context. The tasks that participants were asked to complete are tasks that were constrained to specific domains, primarily mental rotation for the BRT.

Recommendations

- It is strongly recommended that an item analysis be conducted of the BRT. Such an analysis can produce results indicating what items are more difficult than others. It could also indicate differential chronometric models based on block array complexity or angular disparity.
- Complete an incremental validity study to determine if and to what degree the BRT-Tracking combination adds explanatory power to flight training performance to that of the ASTB.
- Produce “cleaner” performance models for the BRT and Tracking tasks by means of gathering reliability data and determine if performance models (e.g., ideal vs. less than ideal) can be differentially classified for the BRT.

- Determine contribution of psychomotor ability to BRT and Tracking tasks via comparison by using “purer” psychomotor tasks (e.g., finger tapping).
- Determine the utility of these tasks outside of U.S. Naval Aviation selection (e.g., diving community), and/or outside of the personnel selection arena (e.g., brain injury diagnosis and therapy).
- Conduct a factor analysis study to confirm the structure of hypothesis variables.

Conclusion

These findings could ultimately result in the utilization of a performance-based measure for selection U.S. Naval Aviation personnel for training. It was found that the experimental tasks, although not entirely in line with hypothetical reasoning, could constitute a good computer-based measure of spatial ability. The research conducted here provides solid evidence in support of the notion that computerized, performance-based assessments can reliably quantify constructs that are similar to those that are currently measured by paper-and-pencil tests. If support is eventually found in favor of increased ecological and predictive validity of the novel tasks introduced here, it will result in a significant, demonstrable savings in training dollars and possibly relate to increased Navy aviation safety.

A more general significant finding in this study was the performance algorithm of the BRT, similar to that of Shepard and Metzler (1971), where a strong linear relationship was found between response time and degrees of separation among mentally rotated stimuli. Shepard and Metzler also found that, on average, mental rotation takes place at the rate of 60° per second. Although not presented in terms of degrees, a similar human performance model was presented here and may be of some value to general knowledge with regard to human ability.

REFERENCES

- Ackerman, P.L. (2001). New approaches to assessment and evaluation of perceptual speed abilities. *Georgia Institute of Technology School of Psychology Technical Report DTIC ADA419862*: Atlanta.
- Ackerman, P.L. & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology*, 78, 413-432.
- Ackerman, P.L. & Lohman, D.F. (1990). An investigation of the effect of practice on the validity of spatial tests. *Final Report (NPRDC Contract N66001-88C-0291)*. Minneapolis, MN: Personnel Decisions Research Institute.
- Amorim, M.A, Isableu, B., & Jarraya, M. (2006). Embodied spatial transformations: "Body analogy" for the mental rotation of objects. *Journal of Experimental Psychology: General*, 135, 327-347.
- ASTB. (2006). Retrieved March 3, 2006 from <http://www.usnavy.vt.edu/CurrentStudents/Forms/astboverview.pdf>
- Binet, A. & Simon, T. (1916). *The Development of Intelligence in Children* (E.S. Kite, Trans.). Baltimore: Williams and Wilkens.
- Bowers, C. A., Milham, L. M., & Price, C. (1998). Dual-task results and the lateralization of spatial orientation: Artifact of test selection? *Journal of General Psychology*, 125(1), 5-16.
- Buchner, A., Erdfelder, E. & Faul, F. (2001). G-Power 2.0. Retrieved from <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/> on 6 June 2008.
- Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation, Space, and Environmental Medicine*, 60, 46-49.
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Cambridge, England: Cambridge University Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Delaney, H.D. (1990). Validation of dichotic listening and psychomotor task performance as predictors of primary flight training criteria: Highlighting relevant statistical issues. *Technical Report No. AD-A227 010*, Naval Aerospace Medical Research Laboratory.
- Fleishman, E.A. (1956). Psychomotor selection tests: Research and application in the united states air force. *Personel Psychology*, 9, 449-467.
- Geary, D. C., & DeSoto, M., C. (2001). Sex differences in spatial abilities among adults from the United States and China: Implications for evolutionary theory. *Evolution and Cognition*, 7, 172-177.
- Ghez, C. & Krakauer, J. (2000). The organization of movement. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 653-673). New York: McGraw-Hill.
- Gibson, J.J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology*, 49, 182-194.
- Griffin, G.R. (1987). Development and evaluation of an automated series of single-and multiple-dichotic listening and psychomotor tasks. *Technical Report No. AD-A199490*, Naval Aerospace Medical Research Laboratory.
- Guilford, J.P. (1958). A system of the psychomotor abilities. *American Journal of*

- Psychology*, 71(1), 164-174.
- Guilford, J.P. & Lacey, J.I. (Eds) (1947). Printed classification tests. *AAF Aviation Psychology Research Program Reports (No. 5)*. Washinton D.C.: GPO.
- Guilford, J. P., & Zimmerman, W. S. (1948). The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology*, 32(1), 24-35.
- Johnson A.M. (1990). Speed of mental rotation as a function of problem solving strategies. *Perceptual and Motor Skills*, 71, 803-806.
- Johnston, P.J. & Catano, V.M. (2002). Psychomotor abilities tests as predictors of training performance. *Canadian Journal of Behavioural Science*, 34, 75-83.
- Kandel, E.R., Kupferman, I., & Iversen, S. (2000). Learning and Memory. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 1227-1245). New York: McGraw-Hill.
- Kandel, E.R. & Wurtz, R.H. (2000). Constructing the visual image. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 492-506). New York: McGraw-Hill.
- Kass, S.J., Ahlers, R.H., & Dugger, M. (1998). Eliminating gender differences through practice in an applied visual spatial task. *Human Performance*, 11(4), 337-349.
- Lane, N.E. & Kennedy, R.S. (1990). Users manual for the automated performance test system (APTS). *NASA-CR-185631*. Houston, TX: NASA Johnson Space Center.
- Larson, G.E. (1996). Mental rotation of static and dynamic figures. *Perception & Psychophysics*, 58, 153-159.
- Lee, D.N. (2006). How movement is guided. Retrieved August 20, 2008 from <http://www.perception-in-action.ed.ac.uk/publications.htm>
- Levin, S.L., Mohamed, F.B., & Platek, S.M. (2005). Common ground for spatial cognition? A behavioral and fMRI study of sex differences in mental rotation and spatial working memory. *Evolutionary Psychology*, 3, 227-254.
- Lohman, D.F. (2000). Complex information processing and intelligence. In R.J. Sternberg (Ed.), *Handbook of Intelligence*. Cambridge: Cambridge University Press.
- Lohman, D.F. (1993, July). *Spatial Ability and G*. Paper presented at the first Spearman Seminar, University of Plymouth, UK.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General Intelligence, Objectively Determined and Measured." *Journal of Personality and Social Psychology*, 86, 96-111.
- Merchant, H. & Georgopoulos, A.P. (2006). Neurophysiology of perceptual and motor aspects of interception. *Journal of Neurophysiology*, 95, 1-13.
- Naval Operational Medicine Institute. (2004). APEX.NET examiner's guide (NOMI P5098-C Part II). Pensacola, FL: Naval Operational Medicine Institute.
- Pellegrino, J., Hunt, E., Abate, R., & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behavior Research Methods, Instruments, and Computers*, 19, 231-236.
- Qu, S. (2003). Development and Testing of an Advanced Terrain Awareness and Warning System. Accessed 10 Dec 2006 from: http://ocw.mit.edu/NR/rdonlyres/Aeronautics-and-Astronautics/16-622Fall2003/B8DD3827-99D6-4C06-94B7-73401A5CD1AF/0/qu_shen.pdf
- Saper, C.B., Iversen, S. & Frackowiak, R. (2000). Integration of sensory and motor

- function: The association areas of the cerebral cortex and the cognitive capabilities of the brain. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 1227-1245). New York: McGraw-Hill.
- Shea, D.L., Lubinski, D., & Benbow, C.P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93, 604-614.
- Shepard, S. & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 3-11.
- Shepard, R.N. & Metzler, D. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Snow, R.E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, 3/4, 536-560.
- Sternberg, R.J. (2000). *Handbook of Intelligence*. Cambridge, U.K.: Cambridge University Press.
- Thurstone, L.L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- U.S. Navy Flight Surgeon's Manual. (1989). Retrieved April 2007 from: http://www.iiimef.usmc.mil/medical/FMF/FMFE/FMFEref/fs_man/CHAPTER%2012.html
- Vandenberg, S.G. & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.